Exploring the ability of LSTM-based hydrological models to simulate streamflow time series for flood frequency analysis

We would like to thank the reviewer for their valuable and constructive feedback. We appreciate the time and effort that was put into the review. All concerns have been carefully addressed. Detailed responses to each of the reviewer's comments are presented below. For clarity, the reviewer's comments are presented in black font, with our responses in blue.

Sincerely,

Jean-Luc Martel, on behalf of all authors.

Reviewer 1: https://doi.org/10.5194/egusphere-2024-2134-RC1

The manuscript entitled "Exploring the ability of LSTM-based hydrological models to simulate streamflow time series for flood frequency analysis" presents an interesting comparison between a distributed hydrological model (HYDROTEL) and Long Short-Term Memory (LSTM) deep learning models. Below are some points regarding its methodology, results, and potential areas for improvement:

Thank you very much for your positive comments and suggestions. Please refer to the point-by-point responses to your comments below.

1. LSTM is one class of machine learning algorithms. There are other types being used with good quality of results such as Convolutional Neural Networks (CNNs), Random Forests, or Gradient Boosted Trees. This should be considered in the literature review and/or as a future development.

While we focus this paper on LSTM-based hydrological models, we agree that other types of machine learning algorithms could be used for hydrological modeling, without necessarily outperforming either LSTM-based models or traditional models. We propose to expand the literature review with papers on the proposed topics and more if applicable.

2. One of the key methods tested, oversampling of extreme peak streamflow events, performed poorly. This suggests a more nuanced approach to data augmentation might be required. Future work could explore advanced synthetic data generation techniques like the Synthetic Minority Over-sampling Technique (SMOTE) rather than simply replicating extreme events. One example is the paper: Wu, Yirui, Yukai Ding, and Jun Feng. "SMOTE-Boost-based sparse Bayesian model for flood prediction." EURASIP Journal on Wireless Communications and Networking 2020 (2020): 1-12.

Thank you for this proposition. Indeed, we were expecting oversampling to perform better than it did in the paper. While we already mentioned in the discussion section (Section 4.1 Strengths and weaknesses of each model in streamflow simulation) the possibility to use techniques such as SMOTE, we propose to further elaborate on how future work could address this issue and potentially benefit from oversampling methodologies.

3. The multihead attention mechanism did not significantly improve the LSTM model's performance. This raises questions about whether it was fully optimized or if a different attention configuration could be more effective. The complexity added by the attention mechanism might not have been justified, given the size of the dataset. I know that the codes were shared, but some diagram and/or a more complete description of the attention mechanism would be interesting to be added, to help future research in the area.

Thank you for this suggestion. We propose to add a diagram in the Methods Section (Section 2.4.3 Multihead attention) to provide a complete description of how the multihead attention mechanism was implemented in this study.

4. One of the paper's recurring challenges is the inherent scarcity of extreme flood events, which makes it difficult for LSTMs to train effectively. Although the study attempts to mitigate this issue, it highlights that LSTMs struggle with rare event prediction without sufficient data. The paper could benefit from exploring more advanced techniques for handling imbalanced datasets, such as ensemble methods or using generative models to simulate extreme events.

The prediction of extreme flood events is indeed a challenge when it comes to machine learning algorithms such as the LSTM-based hydrological models used in this study. The goal of this study was to explore the ability of these models to properly simulate streamflow time series that could ultimately be used for flood frequency analyses, which we believe we have managed to do, as highlighted by our results. While some techniques such as the multihead attention mechanism and the oversampling performed poorly, other methods such as the additional donors and the inclusion of traditional hydrological model simulations performed remarkably well. We believe that this is where the largest potential for LSTM-based models resides. By including a larger number of donors and, also, traditional hydrological model simulations on these donors, it is expected that significant gains can be made on the prediction of extreme flood events. Note that we could not incorporate the combination of donor and traditional hydrological models simulations in this study due to the amount of time and computing resources that would have been needed to calibrate the HYDROTEL distributed model on all the donors. However, this could be achieved in future studies by using a simpler model (or an ensemble of models), such as conceptual lumped-based models. Also, as you have suggested, other ensemble methods or generative models could be used to further improve the results.

We propose to expand the discussion (Section 4.4 Should LSTM models be used for peak streamflow simulation?) to further elaborate on this and future works.

5. Given the results across different test periods, there seems to be a risk of overfitting, particularly in models like LSTM-Combined. The paper could benefit from a more thorough discussion and results presentation on the loss function variation during training and testing epochs.

Overfitting is indeed a risk that needs to be addressed when dealing with neural networks, especially a complex LSTM model like the LSTM-Combined version used in our paper. We believe that Figure 2, presenting the results for the loss function over the training, validation and testing periods demonstrates that there is no problematic overfitting in the model. While there is indeed a significant drop in performance between the training and testing periods, the results are still very good, and much better than those obtained from the HYDROTEL distributed hydrological model. However, we agree that we could have expanded the discussion of these results in our paper to highlight this.

We propose to expand the results section (3.1 Training, validation and testing period results) to better highlight the absence of overfitting in the tested model. The danger of overfitting will also be mentioned in the discussion section (Section 4.4 Should LSTM models be used for peak streamflow simulations?).

6. The authors could provide some explanation about the reasons why floods are occurring in Quebec, Canada. Is it increasing the frequency over the years? Are soil or land use reasons for that? Is it related to climate change?

Certainly. In Quebec, there are three different types of mechanisms that lead to a flooding event:

- Snowmelt or a combination of rainfall during the snowmelt period: This is the main mechanism that leads to flood events, especially over larger catchments (>1000 km^2). Freshets typically happen between the months of March and June, leading to one major flood event per year over these catchments. The most extreme flood events occur when there is a combination of synoptic rainfall events over the snowpack with exceptionally warm temperatures. These only occur once per year during the freshet, and so are de facto rare events (proportionally) in the dataset, making it harder to train LSTM-based models on these specific events.
- Synoptic extreme rainfall events or hurricane remnants: These occur mostly on medium- to large-size catchments (approximately between 100 and 1000 km^2), leading to similar or larger runoff volumes that can happen during the snowmelt period. These events can happen multiple times per year.
- Convective extreme rainfall events: This type of flooding event occurs only in very small catchments or urbanized areas, which were excluded from this study.

We propose to add this clarification in the Methods section (Section 2.1 Study area) to provide additional context to the reader with respect to the mechanisms leading to flood events in the study area.

Overall, the paper provides valuable insights into the utility of LSTMs for hydrological modeling, especially in terms of hybrid model approaches.

Thank you very much for your valuable and constructive feedback that helped improve our paper. We hope that our responses answer your comments.